

Phylogenetic Distribution of Genes in Metagenome

From a metagenome detail page (e.g., https://img.jgi.doe.gov/cgi-bin/mer/main.cgi?section=MetaDetail&page=metaDetail&taxon_oid=3300012988), scroll down to find the **Phylogenetic Distribution of Genes** section and click the **Distribution by BLAST percent identities** button (see Figure 1(i)).

(i) [Phylogenetic Distribution of Genes](#)

Distribution by BLAST percent identities

(ii) **Phylogenetic Distribution of Genes in Metagenome**

Genome: [Soil microbial communities amended with fresh organic matter from upstate New York, USA - Whitman soil sample_242_MG](#)

Phylogenetic distribution was computed on 15-JAN-17

Note: gene hit counts may be higher than the actual gene list as some genomes or genes may have been obsoleted recently.

The Phylogenetic Distribution of Genes allows to assess the phylogenetic composition of a genome sample based on the distribution of best BLAST hits of protein-coding genes in the dataset.

[Table View](#) [Tree View](#)

Data Type

Percent Identity
☒ Successive (30% to 59%, 60% to 89%, 90%+)
☐ Cumulative (30%+, 60%+, 90%+) ☐ Display hit genome count

Distribution By
☒ Gene count
☐ Estimated gene copies

Display Options
☒ Show percentage column (only for gene count)
☒ Show histogram column

[Go](#) [Reset](#)

(iii) [Table View](#) [Tree View](#)

View the phylogenetic distribution of genes in a radial tree format.

[Draw Tree](#)

Figure 1. Phylogenetic Distribution of Genes

From the **Table View** tab (Figure 1(ii)), users can make the following selections:

- **Data Type:** assembled, unassembled (if available), or both (if available).
- **Percent Identity:** Successive (30% to 59%, 60% to 89%, 90%+), or Cumulative (30%+, 60%+, 90%+). If the Cumulative option is selected, then users can also select whether to display hit genome count or not.
- **Distribution By:** Gene count, or Estimated gene copies (if available; otherwise gene count will be used)

- **Display Options:** Show percentage column (only for gene count) and/or Show histogram column.

Click the **Go** button to view phylogenetic distribution of genes in this metagenome. The **Reset** button will reset all user selections back to the default selections.

Please note that phylogenetic distribution of genes is pre-computed. The date (highlighted in a green oval in Figure 1(ii)) shows when the pre-computation was done.

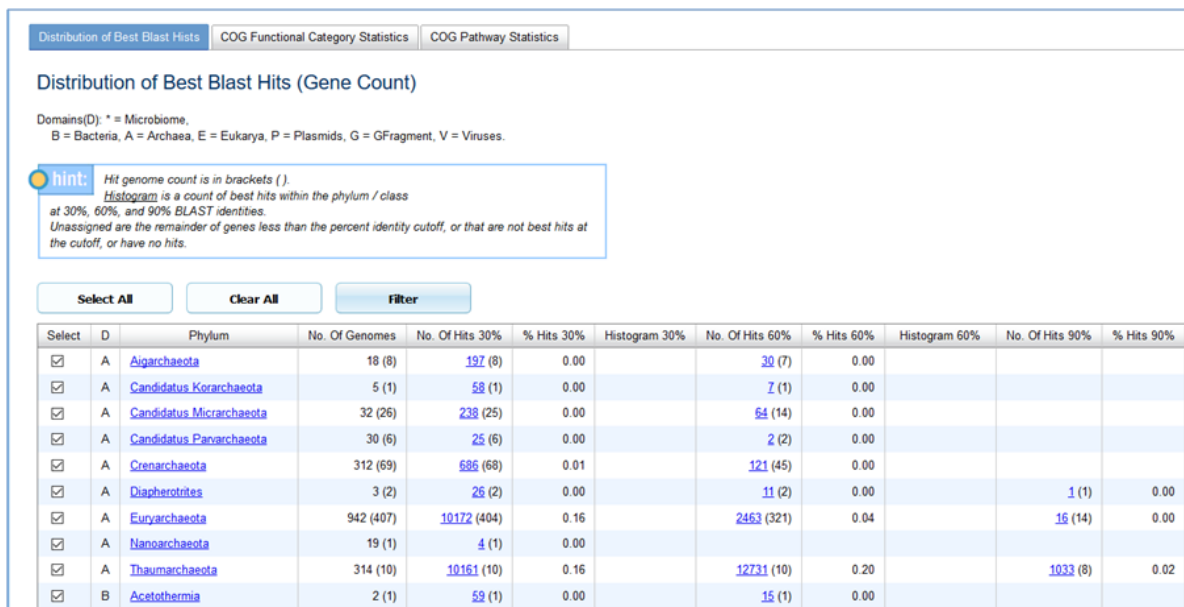


Figure 2. Phylogenetic Distribution of Genes in Metagenome.

Please note that 30%, 60% and 90% correspond to percent identity ranges between the query metagenomic protein and target isolate protein: 30% category includes metagenome proteins for which best hit to isolate proteins has between 30 and 59% identity; 60% category includes proteins for which best hit has 60 to 89% identity to isolate proteins, and 90% category includes metagenome proteins for which best isolate hit has 90+% identity.

IMG uses % identity rather than evaluate filtering of best hits, because evaluates depend on the length of the alignment between query and target proteins. For instance, an evaluate of alignment between two 600 amino acid proteins with ~40% identity would be 1.0e-50. An alignment of nearly identical very short proteins (<50 amino acids) would have similar evaluate of 1.0e-50, even though it is obvious that these short proteins are more similar to each other than the long proteins with the same evaluate. % identity ranges used in IMG roughly correspond to the average amino acid identity found between genomes from the same *genus* (90+% range), the same *family* or *order* (60-89% range), and same *class* or *phylum* (30-59% range).

From the result table in Figure 2, users can click on a phylum to view **Class Statistics** (Figure 3(i)), **Order Statistics** (Figure 3(ii)), all the way down to the *species* level. Users can also click on a gene count to view

the gene list (Figure 3(iii)). Clicking on any gene ID from the list will lead to the corresponding gene detail page.

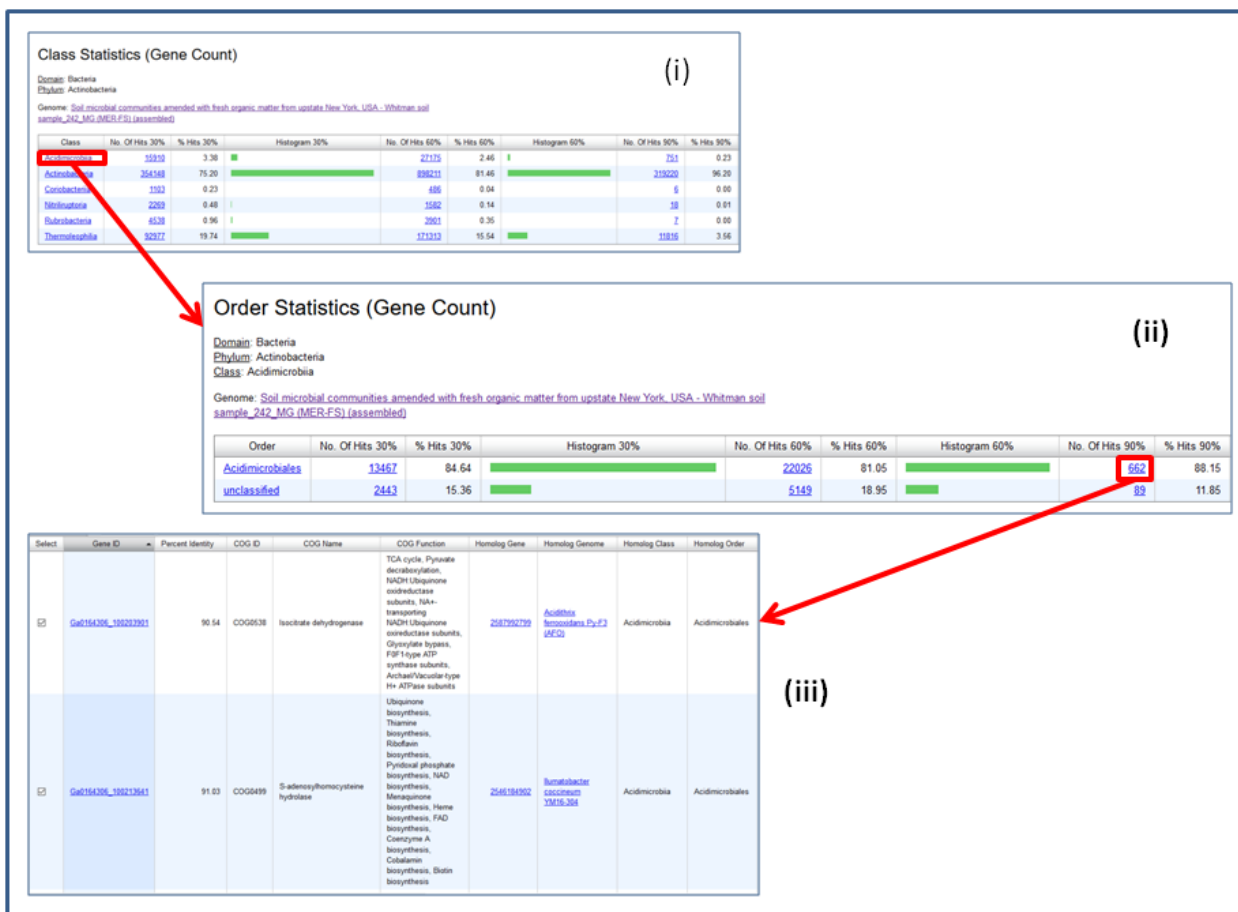


Figure 3. Exploring Phylogenetic Distribution of Genes.

The **COG Functional Category Statistics** tab in Figure 2 provides links to COG functional category information of genes with 30%, 60% and 90% hits, respectively (see Figure 4(i)). Users have the option to view the result in Table format (Figure 4(ii)) or in Chart format (Figure 4(iii)).

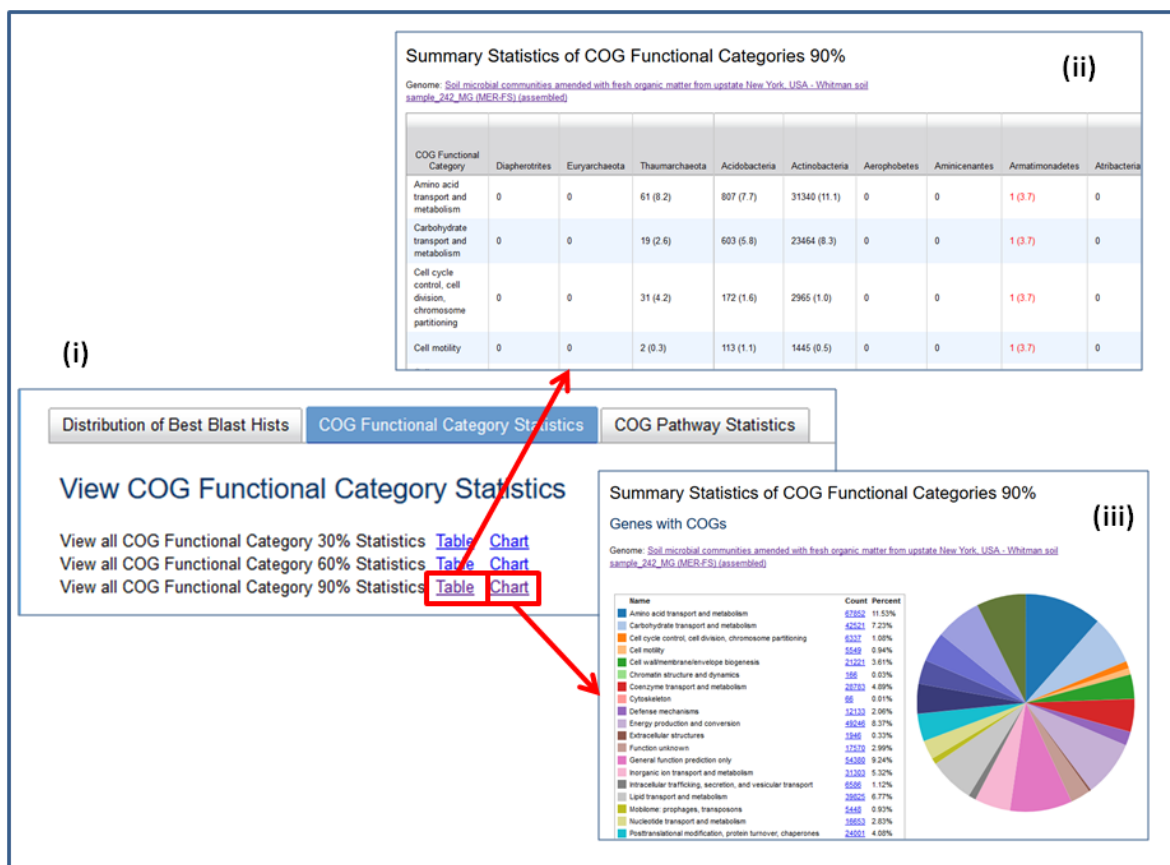


Figure 4. COG Functional Category Statistics.

The **COG Pathway Statistics** tab in Figure 2 provides links to COG pathway information of genes with 30%, 60% and 90% hits, respectively (see Figure 5(i)). Users have the option to view the result in Table format (Figure 5(ii)) or in Chart format (Figure 5(iii)).

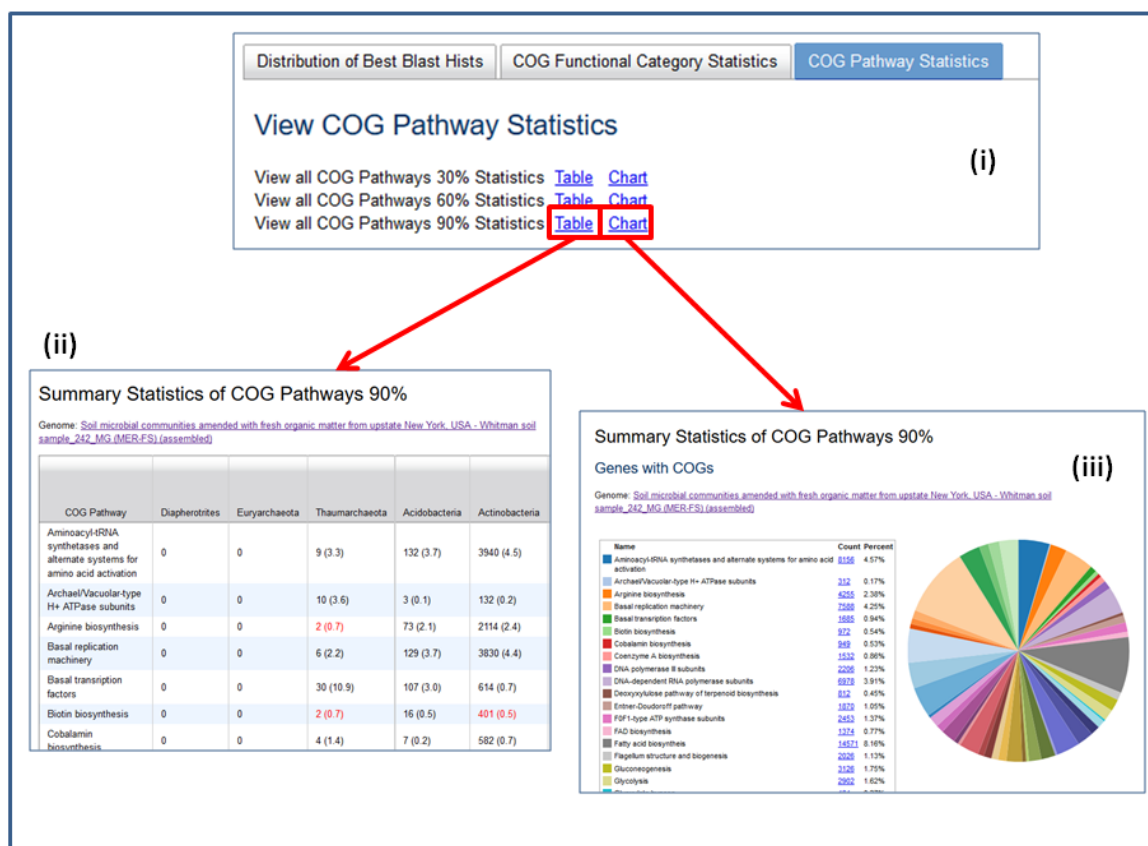


Figure 5. COG Pathway Statistics.

The **Tree View** tab in Figure 1(ii) allows users to view phylogenetic distribution result in a radial tree format. Click the **Draw Tree** button in Figure 1(iii) to view the radial tree display. There is a **Customize Tree** button for users to customize various display parameters such as taxonomic rank to view and tree size (see Figure 6). Users can also export the tree image in PNG, GIFF, JPEG, PDF, PostScript or TIFF format, or export the tree data. Click the **Export** tab to find these exporting functions.

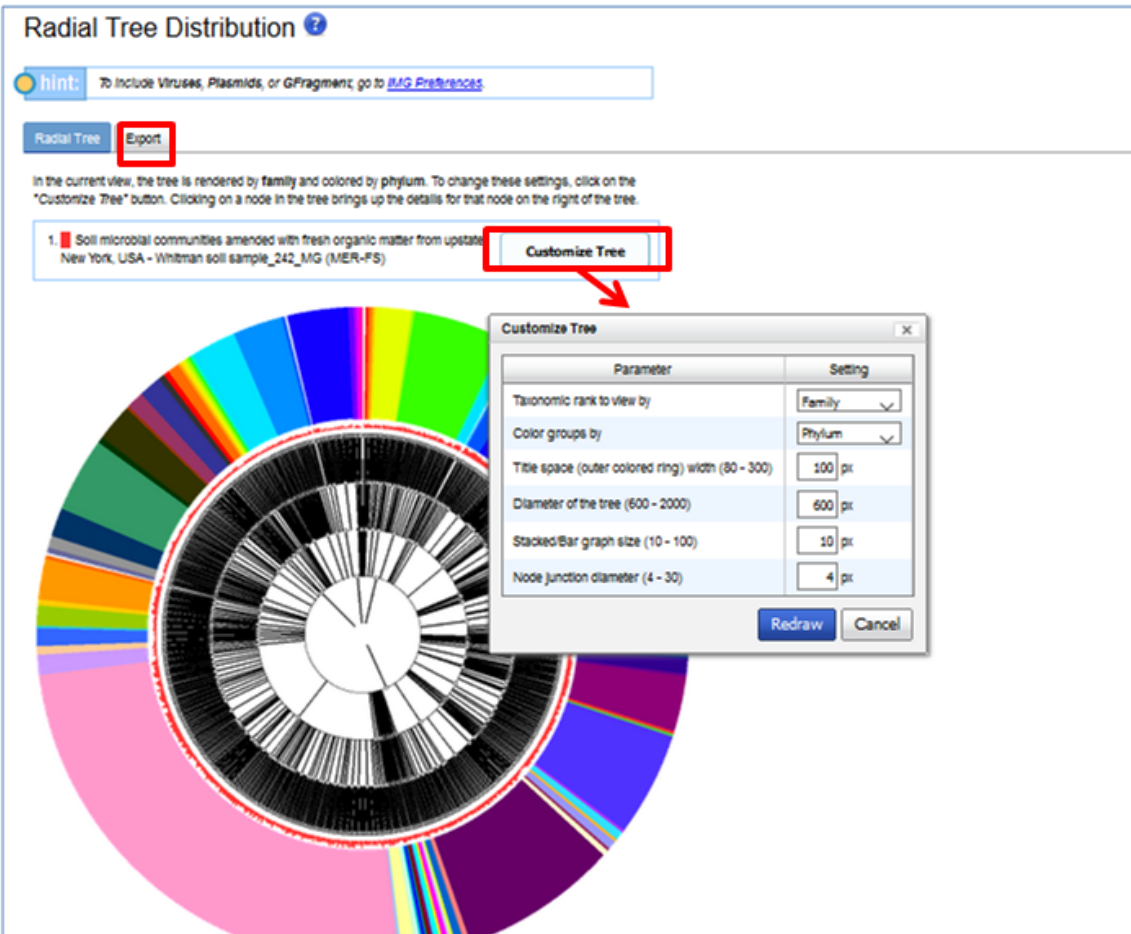


Figure 6. Phylogenetic Distribution of Genes in Radial Tree Display.